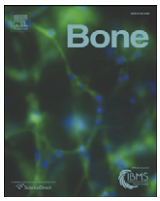
Bone 56 (2013) 16–22

Contents lists available at SciVerse ScienceDirect

Bone

journal homepage: www.elsevier.com/locate/bone



Original Full Length Article

Comparison of different screening tools (FRAX®, OST, ORAI, OSIRIS, SCORE and age alone) to identify women with increased risk of fracture. A population-based prospective study

Katrine Hass Rubin [a,b,*], Bo Abrahamsen [a,c], Teresa Friis-Holmberg [d], Jacob V.B. Hjelmborg [e], Mickael Bech [f], Anne Pernille Hermann [b], Reinhard Barkmann [g], Claus C. Glüer [g], Kim Brixen [a,b]

[a] Institute of Clinical Research, University of Southern Denmark, Denmark
[b] Department of Medical Endocrinology, Odense University Hospital, Denmark
[c] Department of Medicine F, Copenhagen University Hospital Gentofte, Denmark
[d] National Institute of Public Health, University of Southern Denmark, Copenhagen, Denmark
[e] Department of Biostatistics, Institute of Public Health, University of Southern Denmark, Odense, Denmark
[f] COHERE, Department of Business and Economics, University of Southern Denmark, Denmark
[g] Section Biomedical Imaging, Department of Diagnostic Radiology, University Hospital Schleswig-Holstein, Kiel, Germany

ARTICLE INFO

Article history:
Received 15 January 2013
Revised 16 April 2013
Accepted 3 May 2013
Available online 10 May 2013

Edited by: Robert Recker

Keywords:
Epidemiology
Risk factors
Screening tools
Prediction
Fracture

ABSTRACT

Purpose: To compare the power of FRAX® without bone mineral density (BMD) and simpler screening tools (OST, ORAI, OSIRIS, SCORE and age alone) in predicting fractures.
Methods: This study was a prospective, population-based study performed in Denmark comprising 3614 women aged 40–90 years, who returned a questionnaire concerning items on risk factors for osteoporosis. Fracture risk was calculated using the different screening tools (FRAX®, OST, ORAI, OSIRIS and SCORE) for each woman. The women were followed using the Danish National Register registering new major osteoporotic fractures during 3 years, counting only the first fracture per person. Area under the receiver operating characteristic curve (ROC) and statistics and Harrell's index were calculated. Agreement between the tools was calculated by kappa statistics.
Results: A total of 4% of the women experienced a new major osteoporotic fracture during the follow-up period. There were no differences in the area under the curve (AUC) values between FRAX® and the simpler tools; AUC values between 0.703 and 0.722 ($p = 0.86$). Also, Harrell's C values were very similar between the tools. Agreement between the tools was modest.
Conclusion: During 3 years follow-up FRAX® did not perform better in the fracture risk prediction compared with simpler tools such as OST, ORAI, OSIRIS, SCORE or age alone in a screening scenario where BMD was not measured. These findings suggest that simpler models based on fewer risk factors, which would be easier to use in clinical practice by the GP or the patient herself, could just as well as FRAX® be used to identify women with increased risk of fracture.
Summary: Comparison of FRAX® and simpler screening tools (OST, ORAI, OSIRIS, SCORE) in predicting fractures indicate that FRAX® did not perform better in fracture risk prediction compared with the simpler tools or even age alone in a screening scenario without bone mineral density assessment.

Introduction

Fragility fractures associated with osteoporosis are common [1] and impose considerable burdens on the individual [2], increased mortality [3] and add significant costs to the society [4]. Approximately 50% of postmenopausal women and 20% of men older than 50 years will experience a fragility fracture in their remaining lifetime [5]. At present, the majority of men and women at high risk of fracture are not diagnosed or treated [6] and several studies have suggested that the case-finding strategies endorsed in many countries perform less than well [7]. Several tools have been developed to integrate risk factors such as age, low body weight, history of fractures and use of glucocorticoids into a single estimate of fracture risk for an individual. These tools are either aimed at identifying individuals with an increased risk of fractures (with the option to include a BMD result in the risk scoring) or identifying individuals at increased

* Corresponding author at: Medical Endocrinology, Institute of Clinical Research, University of Southern Denmark, Odense University Hospital, Kløvervænget 10, 6th floor, DK-5000 Odense C, Denmark. Fax: +45 66 11 33 71.
E-mail address: Krubin@health.sdu.dk (K.H. Rubin).

8756-3282
http://dx.doi.org/10.1016/j.bone.2013.05.002

risk of having low BMD. However, because the effect of BMD on fracture risk is in itself influenced by the presence of clinical risk factors, fracture risk tools have also been used to guide physicians in whether to refer patients to a BMD measurement or not [8].

Fracture Risk Assessment Tool (FRAX®) uses 10 clinical risk factors and can be used with or without bone mineral density (BMD) to predict the 10-year probability of hip fractures or major osteoporotic fractures in patients (clinical spine, forearm, hip or shoulder fracture) [9,10]. The recently updated National Osteoporosis Foundation (NOF) guidelines recommend treatment of individuals with an increased risk of fracture based on the FRAX® [11]. This involved postmenopausal women and men aged 50 years and older with low bone mass (T-score between −1.0 and −2.5, osteopenia) at the femoral neck or spine and a 10-year hip fracture probability ≥ 3% or a 10-year major osteoporotic fracture probability ≥ 20% as calculated by the FRAX® tool [11]. FRAX® has been validated in 11 independent cohorts [9], and country specific adaptations are available to a large number of countries, including Denmark [9]. Simpler approaches have also been suggested. Age is strongly associated with fracture risk [1] and the U.S. Preventive Services Task Force (USPSTF) recommends screening with DXA in all women aged 65 years and older and in women below 65 years with increased risk of fracture (whose 10-year fracture risk is equal to or greater than that of 65-year-old white women without additional risk factors; 9.3% based on FRAX® calculation); diagnosis and treatment are determined from DXA result [12]. NOF also recommends DXA testing in women above 65 years and women aged 50–65 years with high risk factor profile [11].

BMD has also a strong association with fracture risk where individuals with low BMD have progressively higher risk of fracture [13]. Several tools based on fewer clinical risk factors are available to predict low BMD. As discussed above, the justification for such tools is primarily to identify women who are more likely to have low BMD and then could undergo BMD measurement for a definitive assessment. The simplest tool is the Osteoporosis Self-assessment Tool (OST), which is based on age and body weight alone [14], while others include more risk factors in addition to age and weight: the Osteoporosis Risk Assessment Instrument (ORAI) [15], the OSteoporosis Index of RISk (OSIRIS) [16], and the Simple Calculated Osteoporosis Risk Estimation (SCORE) [17]. All these tools have been developed in women, validated in independent cohorts and the performance of the tools was similar to that seen in the development cohorts [15,18–20]. OST has been validated in both men [21] and women [20,22]; validation studies of the other tools included only women.

Since the release of FRAX® in 2008, a number of studies have compared the performance of FRAX® with other online risk algorithms with an outcome of 5 or 10-year probability of fractures and several other parsimonious models including age. Most of these studies conclude that simpler models perform as well as FRAX® in predicting fractures. Kanis et al. [23] have criticized the conclusions of these studies in part because of the comparison of FRAX® with what Kanis et al. called “home grown” models. Such bespoke models included age or BMI alone, age plus BMI, age plus previous fracture. OST, ORAI, OSIRIS and SCORE include some of the same risk factors and they are also simpler than FRAX. However, tools will always perform well within the derivation cohort and the test of their performance lies in verification within other cohorts.

To date none has tested the performance of FRAX® compared with the simple well validated osteoporosis risk assessment tools (ORAI, OSIRIS, OST and SCORE) and it is uncertain whether FRAX® performs better that these simpler tools. Therefore the aim of the present study was to compare the power of FRAX® (without BMD) and simpler screening tools (OST, ORAI, OSIRIS, SCORE and age alone) in predicting fractures. We hypothesized that the more complex FRAX® (without BMD) tool predicts fracture better than OST, ORAI, OSIRIS, SCORE and age alone.

Methods

Design

This study was a prospective, population-based study performed in the Region of Southern Denmark. Study design and baseline data have been reported previously [24]. In brief, data on self-reported risk factors were collected in a random sample of the population in spring 2009. Data regarding fractures (type and date) during follow-up were extracted from the Danish National Patient Register (NPR) and information on death and emigration were extracted from the Danish National Civil Registration System (NCR) after three years of follow up.

Study population

From the NCR we randomly selected 5000 women living in the Region of Southern Denmark, aged 40–90 years, stratified by decades. During the period from March to May 2009, a self-administered questionnaire concerning risk factors for osteoporosis was issued to the study population together with a pre-paid return envelope. Reminders were mailed to non-respondents twice. All women returning a questionnaire were included in the analysis, with the exception of those diagnosed with and treated for osteoporosis. Signed and returned questionnaires were considered as informed consent to be included in the analysis. All participants were anonymized and the study was approved by the Local Ethical Committee.

Questionnaire

The questionnaire was designed to enable calculation of fracture risk based on each tool at an individual level. It therefore comprised items on weight, height, ethnicity, history of osteoporosis, personal and family history of fracture, smoking habits, consumption of alcohol, use of oral glucocorticoids, use of oestrogen, and diseases associated with secondary osteoporosis (*e.g.* rheumatoid arthritis, type 1 diabetes, osteogenesis imperfecta, untreated long-standing hyperthyroidism, hypogonadism or premature menopause (<45 years), chronic malnutrition, or malabsorption and chronic liver disease). The questions were constructed to allow answering by simple “yes”, “no” or “don't know”, however, body height and weight could be entered as digits. The questionnaire was validated and the reliability tested as previously reported [24]. The questionnaire was read by optical character recognition (OCR); the accuracy of this setup was previously tested without any difference in data registration [24].

Fracture risk prediction tools

Self-reported baseline data were used to calculate the 10-year probability of fracture by FRAX® and to calculate the risk estimate using the simpler tools, ORAI, OSIRIS, OST and SCORE in each woman. Further, age alone was used in the analysis, where the age of the women is used as a simple continuous variable. The number of risk factors used in each tool varies from two in OST to 10 in FRAX®. Table 1 shows the clinical risk factors included in each tool.

Since the detailed algorithm for FRAX® is still not in the public domain, the 10-year probability of fracture was calculated by individual risk scoring using the Danish version of FRAX® [25] using a call of the FRAX® website (version 3.4) [26]. ORAI, OSIRIS, OST and SCORE are instruments designed to predict low BMD. The scoring system for ORAI [15] is as follows: +2 points for non-current usage of estrogen; +9 points for a body weight of less than 60 kg or +3 points for a body weight between 60 and 70 kg and 0 points for weight above 70 kg; and +15 points for ages 75 years or more, +9 points for ages between 65 and 74 years, +5 points for ages between 55 and 64, and 0 points for ages between 45 and 54. To calculate the OST score [14], age was subtracted from weight, the result multiplied by

Table 1
Clinical risk factors included in each tool.

	FRAX	SCORE	OSIRIS	ORAI	OST
Age	X	X	X	X	X
Weight	X	X	X	X	X
Previous low energy fractures	X	X	X		
Estrogen therapy		X	X	X	
Rheumatoid arthritis	X	X			
Height	X				
Parental hip fractures	X				
Smoking	X				
Alcohol	X				
Glucocorticoid therapy	X				
Secondary osteoporosis	X				
Sex	X				
Ethnicity		X			

0.2 and truncated to yield an integer. The OSIRIS score [16] was calculated by adding the index value weighted for each variable: weight (kg) × 2 and remove last digit; age (year) × −2 and remove last digit; +2 if a current HRT user, and −2 if the women have a history of low impact fracture. The SCORE index [17] was calculated as: +5 points for a race other than black; +4 points for rheumatoid arthritis sufferers; +4 points for non-traumatic fractures (wrist, hip and rib) over the age of 45 years; up to a total of 12 points; +1 if the patient never used HRT, 3 times the first digit of the patient's age, and −1 times body weight in pounds divided by ten and truncated to an integer.

In the analyses, we primarily used the nominal score from each tool. In analyses with tools divided into high and low risk of fractures the following dichotomous cut-offs were used: <2 for OST, ≥6 for SCORE, ≥9 for ORAI, ≤1 for OSIRIS, and ≥20% for FRAX® (probability of major osteoporotic fractures). These cut-offs are based on the suggestion of their developers and from validation studies of the tools in Caucasian populations [11,15,19,22,27].

Follow-up

Incident fracture outcomes for this analysis included "major osteoporotic fractures" (FRAX®-defined major osteoporotic fracture; hip, clinical vertebral, wrist or humerus fracture) (ICD-10 codes: S120, S121, S122, S220, S221, S320, T08, S422, S423, S720, S721, S722, S525, S526), and any "osteoporotic fractures" (all fractures except fractures of fingers, toes, skull or face) (ICD-10 codes: S12, S22, S32, S42, S52, S72, S82, T08) during the follow up period. Fracture information on the 5000 women was collected from NPR in April 2012. This register covers all in- and out-patient records in Danish hospitals. Since all persons in Denmark are assigned with a unique personal identification number at birth, it is possible to link data from all public registers at an individual level [28]. Records are available for any given International Classification of Diseases code and surgical procedure [29]. The register has a high validity also regarding the diagnosis of fractures [30,31]. Fractures during the follow-up were counted conservatively as the first fracture (in each category) in each person to avoid overestimating rates due to readmissions. Hip fracture entries with no appropriate surgical code associated were excluded [32]. Follow-up information on death and emigration was also collected in April 2012.

Statistical analysis

Data are shown as mean ± SD or median (range) as appropriate. Frequency tables are used to present the prevalence of each risk factor. Chi-square test (2-sided) for categorical variables and t-test for continued variables were applied to test the difference in baseline characteristics of women with and without fractures during follow up. p-values below 0.05 were considered statistically significant.

Kaplan–Meier curves of cumulative incidence of major osteoporotic fractures are shown for three years of follow-up divided in high and low risk of fractures in the different tools and age alone. Competing risk regressions as alternative to the Kaplan–Meier curves were conducted with incident fractures and death as failure. This analysis was compared to the Kaplan–Meier results to assess the influence of censorings not independent of occurrence of fractures.

We used receiver operating characteristic (ROC) curve analyses to assess the ability of each tool to discriminate between women with or without incident fractures. The AUC of each risk assessor for fracture at follow-up was modeled by univariate logistic regression on the risk assessor as only explanatory variable. In order to adjust for censored women and take time to event (fracture) into consideration, we estimated the Harrell's C index by Cox regression modeling. Harrell's C is analog to AUC in a survival setting. Standard errors robust for cross validation were achieved by the Jack knife-method. Tool assessors with AUC statistics of 0.50 do not perform better than chance alone, while tools with higher AUC statistics perform better than chance. We compared AUC statistics between FRAX® and simpler tools using the "roccomp" procedure in STATA. Finally, the population was divided into quartiles based on fracture risk as predicted by each tool and compared the observed fracture rates across the quartiles. Agreement as to how well each tool assigned the women to risk quartiles was tested using weighted kappa statistic. All analyses were conducted using STATA 12.

Results

As previously reported [24], the respondent rate to the questionnaire was 84%. A total of 334 questionnaires were blank or had several missing items and were excluded leaving 3860 complete questionnaires. We further excluded, 246 women diagnosed with and treated for osteoporosis, leaving 3614 women for analysis. The follow-up period ranged from March 2009 to April 2012. Mean follow up time in the total cohort was 36 months (range 30 to 37 months) and the total follow-up comprised 10,385 person-years. During follow-up, 156 (4%) women suffered "major osteoporotic fractures", 225 (6%) women sustained an "osteoporotic fracture", 174 women died and 6 were lost to follow-up.

The Kaplan–Meier plots of cumulative incidence of major osteoporotic fracture are shown in Fig. 1. The 3 year cumulative "major osteoporosis fracture" estimates for all the tools were similar and ranged at high risk of fracture from 8% in the FRAX® curve to approximate 6% for the SCORE tool. Nearly identical curves were seen in competing-risks regression (data not shown).

Baseline characteristics of the study population overall and stratified according to incident fractures are shown in Table 2. The mean age of the women was 64 ± 13 years and mean BMI was 26 ± 5 kg/m^2. Women with incident fractures were older (mean age 73 ± 11 versus 63 ± 13 years, $p = 0.001$), had more frequent history of fractures (22% versus 9%, $p < 0.001$) and history of falls during the previous 12 months (14% versus 6%, $p < 0.001$), had diseases more often related to secondary osteoporosis (26% versus 18%, $p = 0.011$), and had less frequently used estrogen currently (3% versus 11%, $p = 0.001$).

ROC curve analysis was used to assess the discrimination between the tools. AUC values were very similar (0.703 to 0.722) with no significant differences ($p = 0.86$) in the AUC values between FRAX® and the more simple tools (Table 3). Also, Harrell's C values were very similar between the tools and identical to the AUC values of the different tools.

Restricting the analysis to women aged 50+ years or 65+ did not change the nonsignificant differences in the AUC values between the tools, only the AUC values were lower; about 0.66 and 0.59, respectively (data not shown).

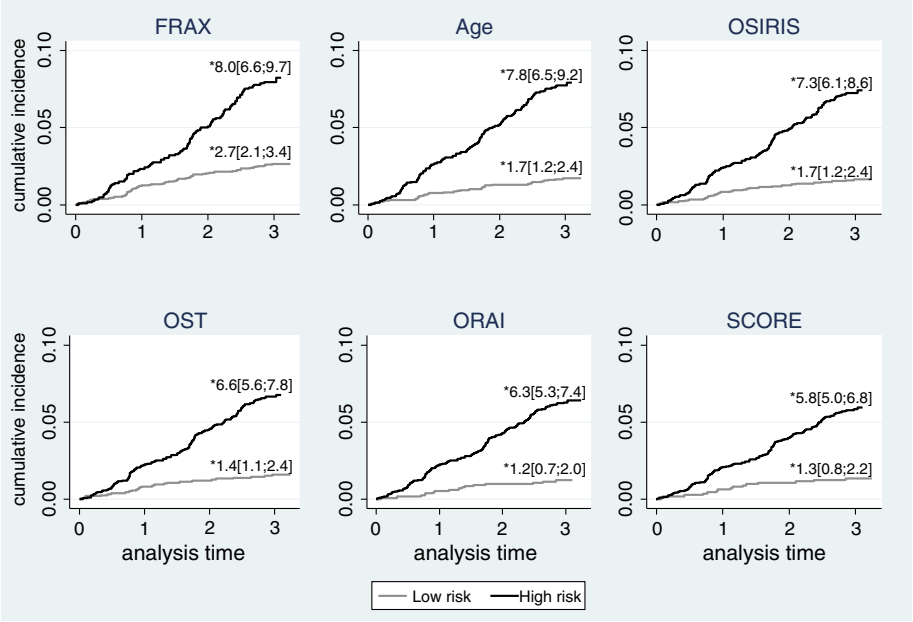


Fig. 1. The Kaplan–Meier plots of cumulative incidence of major osteoporotic fractures divided in high and low risk of fracture for the different tools and for age alone. *3 year fracture estimates. #Division into high and low risks depends of the chosen cut-off for each tool. High risk corresponds to: FRAX® ≥ 20%, age ≥ 65 years, OSIRIS ≤ 1, OST < 2, ORAI ≥ 9, and SCORE ≥ 6.

The observed incidence of fractures in women was plotted against quartiles of predicted risk of fractures from each tool. The tools and age alone performed similarly (Fig. 2). The percentages of women in the highest risk quartile who had a major osteoporotic fracture were approximately 8% for all tools.

Agreement between the tools when assessed using weighted kappa statistic was modest for quartiles of predicted risk of fractures and women with incident fracture. The weighted kappa was best for FRAX® versus age alone (0.73). It was good for FRAX® versus ORAI (0.65) and for FRAX® versus SCORE (0.64), moderate for FRAX® versus OSIRIS (0.53) and for FRAX® versus OST (0.48).

Regarding major osteoporotic fractures, the proportion of women in the highest risk quartile of FRAX®, who also were in the highest quartile for other tools, was 88% for SCORE, 83% for age alone, 79% for ORAI, and 78% for both OST and OSIRIS. Restricting the analysis to women aged 50+ years did not change the results (data not shown).

Table 2
Baseline characteristics of the 3614 women stratified according to incident major osteoporotic fractures.

Variable	All women N = 3614 N (%)	Fracture N = 156 N (%)	No fracture N = 3458 N (%)	p-value*
Age year, mean (SD)	64 ± 13	73 ± 11	63 ± 13	0.001
40–50	694 (19)	3 (2)	691 (20)	p < 0.001
51–60	826 (23)	17 (11)	809 (23)	
61–70	859 (24)	40 (26)	819 (24)	
71–80	680 (19)	41 (26)	639 (19)	
81–90	555 (15)	55 (35)	500 (15)	
BMI kg/m^2, mean (SD)	26 ± 5	25 ± 4	26 ± 5	0.07
<19 kg/m^2	116 (3)	5 (3)	111 (3)	0.62
≥19 kg/m^2	3498 (97)	151 (97)	3347 (97)	
History of parental hip fracture	374 (10)	16 (10)	358 (10)	0.55
History of low energy fracture	337 (9)	35 (22)	302(9)	p < 0.001
Smoking	742 (21)	38 (24)	704 (20)	0.14
Alcohol use, >2 drinks/day	69 (2)	2 (1)	67 (2)	0.42
Rheumatoid arthritis	175 (5)	9 (6)	166 (5)	0.34
Glucocorticoid therapy	131 (4)	7 (5)	124 (4)	0.34
Secondary osteoporosis	655 (18)	40 (26)	615 (18)	0.01
Current estrogen therapy	370 (10)	5 (3)	365 (11)	0.001
Non-black race	3604 (99.7)	156 (100)	3448 (99.7)	0.64
History of falls	218 (6)	22 (14)	196 (6)	p < 0.001

* p value between the group with fractured and nonfractured women.

Discussion

In this study we found that FRAX® and simpler screening tools such as OST, ORAI, OSIRIS, SCORE and even age alone performed similarly in predicting fractures in a screening scenario without BMD assessment. The comparison between tools was based on the AUC and the Harrell's C index by Cox regression modeling and the results were virtually identical for all the tools.

Our results are comparable with the results of several other studies comparing FRAX® both with simple tools and more elaborate tools [33–38]. Most of these studies have included age in the construction of new models. Ensrud et al. [35] included models based on age and BMD or fracture history in comparison with FRAX® in a cohort study of 6652 women with 10-years of follow-up. They concluded that the simple models based on age and BMD or age and fracture history alone predicted the 10-year probability of fractures as well as the more complex FRAX® model. These findings were based on older women (mean age 71 years) and the simple model has not yet been validated in independent populations. Bolland et al. [33] compared age, the Garvan calculator and FRAX® in using data from

Table 3
Area under the curve for receiver operating characteristic curves for each tool and for age alone.

3614 women	Major osteoporotic fractures (N = 156)	All fractures (N = 225)
	AUC (95% CI)	AUC (95% CI)
FRAX	0.722 (0.686, 0.758)	0.701 (0.668, 0.735)
Age alone	0.720 (0.685, 0.755)	0.694 (0.660, 0.727)
ORAI	0.713 (0.677, 0.749)	0.690 (0.658, 0.723)
SCORE	0.712 (0.675, 0.750)	0.691 (0.657, 0.725)
OST	0.704 (0.663, 0.745)	0.682 (0.646, 0.717)
OSIRIS	0.703 (0.664, 0.742)	0.681 (0.646, 0.716)

No significant differences in AUC values were seen between algorithms (p = 0.86 for major osteoporotic fractures and p = 0.56 for all fractures).

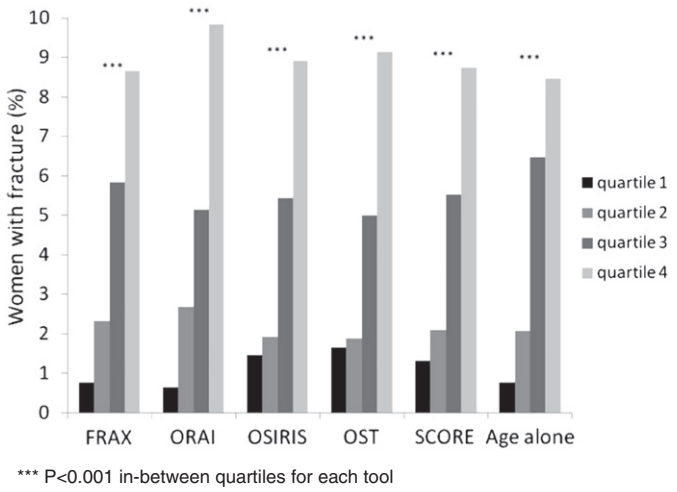


Fig. 2. Observed prevalence of major osteoporotic fractures plotted according to quartiles of predicted fracture risk based as estimated by FRAX®, ORAI, OSIRIS, OST, SCORE tool and age alone. ***p < 0.001 in-between quartiles for each tool.

a RCT regarding calcium supplementation in New Zealand comprising 1422 women aged 55+ years with a follow-up period of 8.8 years. They concluded that FRAX® and the Garvan calculator had moderate discriminative ability for fractures and did not have greater discrimination than simpler models based on age and BMD. This study was also based on older women (mean age 74 years). Incident fractures were recorded by telephone interview and only 57 hip fractures occurred over the 8.8 years of follow-up.

Our results are based on data without BMD using an approach similar to that of the GLOW study [36], which comprised an international cohort with 19,586 women aged 60+ years who had consulted their physician in the past 24 months. They found that a simple model consisting of age and prior fractures performed as well as FRAX® and the Garvan calculator when BMD was unknown. As in our study, they based assessment on self-reported clinical risk factors; however, they used self-reported incident fractures during 2 years of follow-up while we collected fracture data from national registers. We invited participants from a random selection in the general population and had a high responder rate (84%). In contrast, the GLOW study group acknowledged that their sample was prone to bias due to the selection of physicians and due to the sampling and recruitment of patients [36]. Also, their model (with age and prior fracture) was not validated in independent populations. Several other studies have also compared FRAX® with other more elaborate tools such as the QFracture algorithm [34] and the Garvan calculator [33,37] arriving to the same conclusions as the studies mentioned above.

In our study, agreement between the tools with regard to categorizing women into quartiles of risk for major osteoporotic fracture was moderate. However, agreement between the tools in identifying women at the highest quartile of risk for major osteoporotic fracture was high. Approximately 80% of the women classified in the highest risk quartiles by FRAX® were also categorized as highest risk by all the other tools. Sambrook et al. [36] came to a similar conclusion in the GLOW study and our research supports that if women were selected for treatment based on being in the highest quartile of risk, virtually the same women would meet the threshold for treatment regardless of the tool used.

FRAX® is the most complex tool in this study and incorporate 11 risk factors in the algorithm (and may in addition include BMD), whereas the simpler tools only incorporated between 2 and 6 risk factors (Table 1). All the tools included age and BMI. Additional variables did not appear to improve the performance of the tools. Both age and BMI are associated with fracture risk, however, age is the strongest risk factor [1]. Our study also showed that even age alone performed as well as the FRAX® tool without BMD.

Kanis et al. [23] recently discussed potential pitfalls in external validation of FRAX®. Several studies [33,35,38–40] compared the AUC of ROC curves *across studies*. In the present study we compared the AUC of the different predefined tools *within* the same well defined study population. There are limitations to the ability of ROC analysis to discriminate accurately between the performance of predictive models and it is argued that addition of additional variables in models may improve decision-making without materially increasing the AUC, which reflects the diagnostic performance of the models across the range and not just around the point at which critical decisions are made regarding treatment [23]. Because of this, we also undertook analyses where models were compared at relevant clinical intervention threshold (Fig. 1). Kanis et al. [23] also criticized comparison of "home grown" models with the FRAX® tool using the population from which the "home grown" model was derived. This is a relevant concern as the best model to fit a dataset will invariably be a model developed from that particular dataset even if the diagnostic performance may not at all translate to other populations. In our study, we compared the performance of FRAX® and other models to that of age alone. This is a simple epidemiological tabulation of fracture incidence as a function of age and does not constitute a bespoke model to fit the data. Furthermore, OST, ORAI, OSIRIS and SCORE are already well validated simpler tools derived from other cohorts [15,18–20]. Another limitation accurately identified by Kanis et al. [23] is the comparison between predicted and observed outcomes. Since we do not have 10 years of follow-up we look at the observed fractures and compared it with the FRAX® probability of being in risk of fracture. Moreover, we took time-to-event into account by estimating the Harrell's C which did not influence the results. Same results were seen in the GLOW study [36]; these results also showed that AUC values and Harrell's C values were similar for major osteoporotic fractures. Finally, FRAX® adjusts for risk of death while the other tools do not. Our findings, however, were robust to competing-risks regression with both incident fractures and death as failure as alternative to Kaplan–Meier analysis.

In the analyses with each tool dividing participants into those with high versus low risk of fracture we chose to use the cut-off suggested by the developers from validation studies of tools in Caucasian populations. Different cut-offs have been also recommended even among Caucasian populations from studies validating the tools but there was no clear agreement regarding cut-off values for the different tools [41–44]. One study by Rud et al. [41] investigated the performance of SCORE, OST and ORAI in a Danish population. The sensitivity of SCORE, OST and ORAI was 69%, 90% and 50%, respectively, when applied as described by the developers. The authors also tried different cut-offs with higher sensitivities, but since the study only included peri- and early postmenopausal women (mean age 50.5 years) and there are no other studies on Danish women confirming the suggested cut-off from Rud et al. [41] we found it most reasonable to use the cut-offs from the developer of the tools in this study.

The aim of the different tools, i.e. FRAX® with OST, ORAI, OSIRIS or SCORE, differs. FRAX® predicts the probability of fractures while ORAI, OSIRIS, OST and SCORE are designed to predict low BMD. However, since BMD predicts fracture [45] and low BMD are strongly associated with risk of fractures [13] the output of the simpler tools may be perceived as a proxy for probability of fractures. Furthermore, SCORE, OST and ORAI have once each in three different studies been validated with fracture outcome [46–48]. The overall conclusions from these studies were that tools to predict low BMD modestly correlate with clinical fractures.

Other tools such as the Garvan calculator and the QFracture algorithm have similar aim as FRAX®, but we were unable to calculate the fracture risk of these tools since we have no data on the number of falls but only data on whether participants have been falling more

than once the last year. In our study population prior falls were significantly more frequent in fracture cases than in non-fracture cases (14% versus 6%, $p < 0{,}001$).

Our study had a number of important strengths. First, it was a large prospective population-based and including a wide age range (40–90 years). Thus, the results may be applicable to the wider population of women. Second, we had a high response rate and 77% of the invited population were available for analyses. Third, the questionnaire was validated in a large number of women prior to the current study and had a high reliability [24]. Finally, the outcome data relied on data from highly valid Danish national registers and ensured nearly complete follow-up [30,31]. Specifically, the diagnosis of fractures in the NPR has previously been shown to be highly accurate [49].

Our study also has some potential limitations. Follow-up was only three years. However, we took time-to-event into account in our analyses and studies with longer follow-up have showed similar results [33,35,39]. We did not measure BMD in our study. This precluded the possibility to investigate the performance of FRAX® with BMD in comparison with the simpler tools. While we cannot exclude the possibility that FRAX® with BMD would perform better than the simpler tools due to the lack of such data, other studies comparing FRAX® with simpler models including BMD showed that FRAX® with BMD had only a slightly higher AUC than FRAX® without BMD and the simpler models [33,35,38,39]. A further limitation could be that the data on clinical risk factors were self-reported and thus potentially prone to bias. One study demonstrated that a cohort of postmenopausal women over-reported their height by a mean of 2.8 cm and underreported their weight by a mean of 2.1 kg [50]. In our study, the use of self-reported height and weight could result in an over-estimation of the 10-year fracture risk because the BMI might be lower than the real BMI. Also, we cannot completely exclude the possibility that women at high risk of fracture were more motivated to participate in this study. Comparison of respondents and non-respondents revealed some differences as previously reported [24]. Finally, only women participated in the study, thus the results are not generalizable to men.

In conclusion our data indicate that – during medium-term follow-up (3 years) and using self-reported clinical risk factors – more complex tools as FRAX® did not perform better in the fracture risk prediction compared with simpler tools such as OST, ORAI, OSIRIS and SCORE or even age alone in a screening scenario where BMD was not measured. These findings suggest that simpler tools based on fewer risk factors, which would be easier to use in clinical practice by the GP or the patient herself, could just as well as FRAX® be used to identify women with increased risk of fracture and therefore should be referred to a DXA scan.

Acknowledgments

This study is a part of the Risk-stratified Osteoporosis Strategy Evaluation study (ROSE study) which is supported by INTERREG 4A (JNR 08/4177), the Region of Southern Denmark (JNR 08/8133 and JNR 11/5761) and Odense University Hospital.

The funding agencies had no direct role in the conduct of the study; collection, management, analysis and interpretation of the data; and preparation, review or final approval of the manuscript.

Conflicts of interest

None.

References

[1] Sambrook P, Cooper C. Osteoporosis. Lancet 2006;367:2010–8.
[2] Mossey JM, Mutran E, Knott K, Craik R. Determinants of recovery 12 months after hip fracture: the importance of psychosocial factors. Am J Public Health 1989;79:279–86.
[3] Johnell O, Kanis JA, Oden A, Sernbo I, Redlund-Johnell I, Petterson C, et al. Mortality after osteoporotic fractures. Osteoporos Int 2004;15:38–42.
[4] Johnell O. The socioeconomic burden of fractures: today and in the 21st century. Am J Med 1997;103:20S–5S.
[5] Bone health and osteoporosis: a report of the surgeon general. Reports of the Surgeon General. Rockville (MD): Office of the Surgeon General (US); 2004.
[6] Vestergaard P, Rejnmark L, Mosekilde L. Osteoporosis is markedly underdiagnosed: a nationwide study from Denmark. Osteoporos Int 2005;16:134–41.
[7] Rubin KH, Abrahamsen B, Hermann AP, Bech M, Gram J, Brixen K. Prevalence of risk factors for fractures and use of DXA scanning in Danish women. A regional population-based study. Osteoporos Int 2011;22:1401–9.
[8] Compston J, Cooper A, Cooper C, Francis R, Kanis JA, Marsh D, et al. Guidelines for the diagnosis and management of osteoporosis in postmenopausal women and men from the age of 50 years in the UK. Maturitas 2009;62:105–8.
[9] Kanis JA, Oden A, Johnell O, Johansson H, De Laet C, Brown J, et al. The use of clinical risk factors enhances the performance of BMD in the prediction of hip and osteoporotic fractures in men and women. Osteoporos Int 2007;18:1033–46.
[10] Kanis JA, Johnell O, Oden A, Johansson H, McCloskey E. FRAX and the assessment of fracture probability in men and women from the UK. Osteoporos Int 2008;19:385–97.
[11] Osteoporosis Foundation National. Clinician's guide to prevention and treatment of osteoporosis. Washington DC: National Osteoporosis Foundation; 2010.
[12] Nelson HD, Haney EM, Dana T, Bougatsos C, Chou R. Screening for osteoporosis: an update for the U.S. Preventive Services Task Force. Ann Intern Med 2010;153:99–111.
[13] Cummings SR, Melton LJ. Epidemiology and outcomes of osteoporotic fractures. Lancet 2002;359:1761–7.
[14] Koh LK, Sedrine WB, Torralba TP, Kung A, Fujiwara S, Chan SP, et al. A simple tool to identify Asian women at increased risk of osteoporosis. Osteoporos Int 2001;12:699–705.
[15] Cadarette SM, Jaglal SB, Kreiger N, McIsaac WJ, Darlington GA, Tu JV. Development and validation of the Osteoporosis Risk Assessment Instrument to facilitate selection of women for bone densitometry. CMAJ 2000;162:1289–94.
[16] Sedrine WB, Chevallier T, Zegels B, Kvasz A, Micheletti MC, Gelas B, et al. Development and assessment of the Osteoporosis Index of Risk (OSIRIS) to facilitate selection of women for bone densitometry. Gynecol Endocrinol 2002;16:245–50.
[17] Lydick E, Cook K, Turpin J, Melton M, Stine R, Byrnes C. Development and validation of a simple questionnaire to facilitate identification of women likely to have low bone density. Am J Manag Care 1998;4:37–48.
[18] Cadarette SM, Jaglal SB, Murray TM. Validation of the simple calculated osteoporosis risk estimation (SCORE) for patient selection for bone densitometry. Osteoporos Int 1999;10:85–90.
[19] Reginster JY, Ben SW, Viethel P, Micheletti MC, Chevallier T, Audran M. Validation of OSIRIS, a prescreening tool for the identification of women with an increased risk of osteoporosis. Gynecol Endocrinol 2004;18:3–8.
[20] Richy F, Gourlay M, Ross PD, Sen SS, Radican L, De CF, et al. Validation and comparative evaluation of the osteoporosis self-assessment tool (OST) in a Caucasian population from Belgium. QJM 2004;97:39–46.
[21] Adler RA, Tran MT, Petkov VI. Performance of the Osteoporosis Self-assessment Screening Tool for osteoporosis in American men. Mayo Clin Proc 2003;78:723–7.
[22] Geusens P, Hochberg MC, van der Voort DJ, Pols H, van der Klift KM, Siris E, et al. Performance of risk indices for identifying low bone density in postmenopausal women. Mayo Clin Proc 2002;77:629–37.
[23] Kanis JA, Oden A, Johansson H, McCloskey E. Pitfalls in the external validation of FRAX. Osteoporos Int 2012;23:423–31.
[24] Rubin KH, Abrahamsen B, Hermann AP, Bech M, Gram J, Brixen K. Fracture risk assessed by Fracture Risk Assessment Tool (FRAX) compared with fracture risk derived from population fracture rates. Scand J Public Health 2011;39:312–8.
[25] World Health Organisation. WHO Fractures risk assessment tool. http://www.shef.ac.uk/FRAX/tool.jsp?country=29; 2009.
[26] Langtved data A/S. www.langtved.dk; 2009.
[27] Kanis JA, McCloskey EV, Johansson H, Strom O, Borgstrom F, Oden A. Case finding for the management of osteoporosis with FRAX—assessment and intervention thresholds for the UK. Osteoporos Int 2008;19:1395–408.
[28] Thygesen LC, Daasnes C, Thaulow I, Bronnum-Hansen H. Introduction to Danish (nationwide) registers on health and social issues: structure, access, legislation, and archiving. Scand J Public Health 2011;39:12–6.
[29] Andersen TF, Madsen M, Jorgensen J, Mellemkjoer JH, Olsen L. The Danish National Hospital Discharge Register. A valuable source of data for modern health science. Dan Med Bull 1999;46:263–8.
[30] Mosbech J, Jorgensen J, Madsen M, Rostgaard K, Thornberg K, Poulsen TD. The National Patient Registry. Evaluation of data quality. Ugeskr Laeger 1995;157:3741–5.
[31] Thygesen LC, Ersboll AK. Danish population-based registers for public health and health-related welfare research: introduction to the supplement. Scand J Public Health 2011;39:8–10.
[32] Ryg J, Rejnmark L, Overgaard S, Brixen K, Vestergaard P. Hip fracture patients at risk of second hip fracture: a nationwide population-based cohort study of 169,145 cases during 1977–2001. J Bone Miner Res 2009;24:1299–307.
[33] Bolland MJ, Siu AT, Mason BH, Horne AM, Ames RW, Grey AB, et al. Evaluation of the FRAX and Garvan fracture risk calculators in older women. J Bone Miner Res 2011;26:420–7.
[34] Cummins NM, Poku EK, Towler MR, O'Driscoll OM, Ralston SH. Clinical risk factors for osteoporosis in Ireland and the UK: a comparison of FRAX and QFractureScores. Calcif Tissue Int 2011;89:172–7.
[35] Ensrud KE, Lui LY, Taylor BC, Schousboe JT, Donaldson MG, Fink HA, et al. A comparison of prediction models for fractures in older women: is more better? Arch Intern Med 2009;169:2087–94.

[36] Sambrook PN, Flahive J, Hooven FH, Boonen S, Chapurlat R, Lindsay R, et al. Predicting fractures in an international cohort using risk factor algorithms without BMD. J Bone Miner Res 2011;26:2770–7.
[37] Sandhu SK, Nguyen ND, Center JR, Pocock NA, Eisman JA, Nguyen TV. Prognosis of fracture: evaluation of predictive accuracy of the FRAX algorithm and Garvan nomogram. Osteoporos Int 2010;21:863–71.
[38] Sornay-Rendu E, Munoz F, Delmas PD, Chapurlat RD. The FRAX tool in French women: how well does it describe the real incidence of fracture in the OFELY cohort? J Bone Miner Res 2010;25:2101–7.
[39] Hillier TA, Cauley JA, Rizzo JH, Pedula KL, Ensrud KE, Bauer DC, et al. WHO absolute fracture risk models (FRAX): do clinical risk factors improve fracture prediction in older women without osteoporosis? J Bone Miner Res 2011;26:1774–82.
[40] Tremollieres FA, Pouilles JM, Drewniak N, Laparra J, Ribot CA, Dargent-Molina P. Fracture risk prediction using BMD and clinical risk factors in early postmenopausal women: sensitivity of the WHO FRAX tool. J Bone Miner Res 2010;25:1002–9.
[41] Rud B, Jensen JE, Mosekilde L, Nielsen SP, Hilden J, Abrahamsen B. Performance of four clinical screening tools to select peri- and early postmenopausal women for dual X-ray absorptiometry. Osteoporos Int 2005;16:764–72.
[42] Cook RB, Collins D, Tucker J, Zioupos P. Comparison of questionnaire and quantitative ultrasound techniques as screening tools for DXA. Osteoporos Int 2005;16:1565–75.
[43] Gourlay ML, Miller WC, Richy F, Garrett JM, Hanson LC, Reginster JY. Performance of osteoporosis risk assessment tools in postmenopausal women aged 45–64 years. Osteoporos Int 2005;16:921–7.
[44] Gourlay ML, Powers JM, Lui LY, Ensrud KE. Clinical performance of osteoporosis risk assessment tools in women aged 67 years and older. Osteoporos Int 2008;19:1175–83.
[45] Marshall D, Johnell O, Wedel H. Meta-analysis of how well measures of bone mineral density predict occurrence of osteoporotic fractures. BMJ 1996;312:1254–9.
[46] Morin S, Tsang JF, Leslie WD. Weight and body mass index predict bone mineral density and fractures in women aged 40 to 59 years. Osteoporos Int 2009;20:363–70.
[47] Wei GS, Jackson JL. Postmenopausal bone density referral decision rules: correlation with clinical fractures. Mil Med 2004;169:1000–4.
[48] Lacroix AZ, Buist DS, Brenneman SK, Abbott III TA. Evaluation of three population-based strategies for fracture prevention: results of the osteoporosis population-based risk assessment (OPRA) trial. Med Care 2005;43:293–302.
[49] Vestergaard P, Mosekilde L. Fracture risk in patients with celiac disease, Crohn's disease, and ulcerative colitis: a nationwide follow-up study of 16,416 patients in Denmark. Am J Epidemiol 2002;156:1–10.
[50] Bridges MJ, Ruddick S. Can self-reported height and weight be used to calculate 10 year risk of osteoporotic fracture? J Nutr Health Aging 2010;14:611–3.